Prediction of Useful Votes on Yelp Reviews

Christina Pan (capan@stanford.edu), John Poothokaran (johnmp@stanford.edu) | Stanford University CS 230

Motivation

Yelp ratings and recommendations help users to search for different restaurants and pick restaurants to visit.

However, not all Yelp reviews are useful, and identifying the most helpful reviews would add to the customer experience and prevent the website from being overly cluttered.

Objective

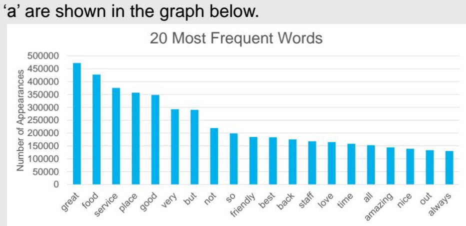
This project aims to build a model that predict the number of useful votes a review should receive based on an analysis of the text, using state of the art natural language processing (NLP) and machine learning (ML) techniques.

Dataset

The dataset consists of the text of reviews gleamed from Yelp's Challenge dataset. It consists of a total of 1.5 million reviews, where the test set has 1.35 M reviews, the dev set has 75k, and test set 75k reviews. All reviews contain at most 50 words for memory reasons.

75% of reviews have 0 useful votes, 15% have 1 vote, and the maximum number of votes is 1341.

This set of reviews contains a vocabulary of 170k words. The 20 most frequent words – aside from words like 'the' or 'a' are shown in the graph below.



Approach

We decided to pursue a variety of NLP architectures of increasingly complexity and assess their performances to find our most optimal model.

All models receive the review's text converted into a word embedding representation (initialized with the GloVe vectors) and output the number of predicted votes. In addition, all models used in the experiments use the RSME loss function and an Adam Optimizer.

Algorithms

Dense

The baseline model consists of an embedding lookup layer, followed by a simple, dense ReLu layer that outputs a prediction for the number of useful votes

LSTM

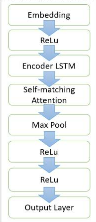
Given the great success and sequential nature of information relayed over reviews, we tried a single layer LSTM.

Modified Bi-attentive Classification Network (BCN)

Since new work in encoder-decoder LSTMs, attention, and CNNs have improved NLP classification tasks, we decided to modify modified the BCN used by McCann et al, with a self matching attention mechanism. (See equations below.)

Due to overfitting and memory issues, the 'integration' LSTM was removed. Below is a diagram showing the layers.

Modified BCN Architecture:



Self-Attention Equations:

$$s_j^t = \mathrm{v}^{\mathrm{T}} \tanh(W_v^P v_j^P + W_v^P v_t^P)$$
$$a_i^t = \exp(s_i^t) / \Sigma_{j=1}^{n} \exp(s_j^t)$$
$$c_t = \Sigma_{i=1}^{n} a_i^t v_i^P$$

Results & Discussion

Results Between Algorithms:

Algorithm	Loss (MSE)	Average Absolute Difference
Dense	0.098	0.313
LSTM	***0.096***	***0.310***
Modified BCN	3.472	1.86

As you can see, the LSTM does the best job on the dev set. Given that the modified BCN is a more complex model, it seems most likely that the modified BCN needs more hyper-parameter tuning.

After conducting qualitative error analysis, we see that the model typically associates strongly opinions as useful ones. In addition, it tends to label reviews where the author writes whether they will come back to a location as useful too. Similarly, the model tends to classify comparative statements (e.g., "Amazing, authentic Yucatan food, not your traditional Tex-Mex Americanized cuisine.") as more useful ones. However, since the model is mostly exposed to restaurant data as a result of Yelp's specialty, it has a hard time finding superlatives in non-restaurant situations.

Future Work

The predictions of the model could be improved by providing other contextual information from the Yelp Dataset and passing the features through different neural architectures. Identifying the usefulness of reviews could be approached as a NLP classification task as well.

Works Cited

- B. McCann, J. Bradbury, C. Xiong, R. Socher. *"Learning in Translation: Contextualized Word Vectors"*. 17th March 2018. URL: https://arxiv.org/pdf/1708.00107.pdf
- Natural Language Computing Group, Microsoft Research Asia. *"R-Net: Machine Reading Comprehension with Self Matching Networks"*. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/05/r-net.pdf